DOCUMENT RESUME

ED 043 024 AL 002 587

TITLE	Proposal for the Development of a Language Testing Program for the Peace Corps.
INSTITUTION	Language Research Foundation, Cambridge, Mass.
SPONS AGENCY	Peace Corps (Dept. of State), Washington, D.C.
PUB DATE	[30 Jun 70]
NOTE	26p.; From the "Final Report on Peace Corps Languae Coordinators Workshop," Rockport, Massachusetts, April 26 to May 16, 1970
EDRS PRICE	EDRS Price MF-$0.25 HC-$1.40
DESCRIPTORS	Evaluation Methods, Feedback, Grammar, *Language Proficiency, *Language Tests, Listening Comprehension, Oral Fxpression, Pronunciation, *Test Construction, *Testing Programs
IDENTIFIERS	*Peace Corps

ABSTRACT

This paper outlines the Educational Testing Service (ETS) plan for developing instruments and procedures for evaluating Peace Corps Trainees' and Peace Corps Volunteers' competence in host country languages at various stages of training or in-country service. The goals of such an evaluation program are first stated. and then a critique of the present evaluation method, which is based on the Foreign Service Institute Interview, is given; deficiences in the areas of listening comprehension, spoken vocabulary, and command of spoken grammar are discussed. Suggestions for changes in the program are made, the desirable specifications for a language evaluation program are outlined, and the program proposed by ETS is described. Final sections deal with the question of the feedback of test information to the student and language staff, and present suggestions for the points in the course of training or service that the tests should be administered. (FWP)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED 043024

PROPOSAL FOR THE DEVELOPMENT OF A LANGUAGE TESTING PROGRAM FOR THE PEACE CORPS

Educational Testing Service

AL 002 587

General Goals of a Comprehensive Language Testing Program

Educational Testing Service believes that its primary contribution to the Peace Corps language training program can be and should be in developing suitable instruments and procedures for evaluating PCT/PCV competence in host country languages at various stages of training or in-country service. There are at least five broad purposes which such an evaluation program would attempt to serve:

1. to provide Peace Corps central administration an indication of the overall improvement (or lack thereof) in the language competence of large groups of Trainees/Volunteers as measured across a reasonably long time span. This would include, for example, periodic comparisons of the average competence of students trained in a given language, or within languages, at various training centers. Provided that language learning variables not associated with the training program as such (e.g. student language aptitude, prior study of the language) are controlled operationally or statistically, general comparisons of this type are valid and should be of interest to Peace Corps administrative groups.

2. to provide more specific feedback to training centers regarding the progress of its trainees toward various well-defined goals of language competence.

3. to provide similar feedback to individual trainees, both for general motivational purposes and to point out specific areas of strength or weakness in their command of the language.

4. to provide the basic language competence data from which, in conjunction with other data collected in the field, it would eventually be possible to determine the nature and extent of language mastery needed for successful work in various in-field activities or job classifications.

5. to facilitate the conducting of certain research studies appropriate to Peace Corps concerns, in particular the question of which aspects of a language should be formally taught during the training period, and which could reasonably be acquired independently by the volunteer in the course of his service.

Critique of Current Program

The present language evaluation program in the Peace Corps relies almost exclusively on the Foreign Service Institute interview [See Appendix I], in which students are rated on a 1 to 5 scale following a fifteen- or twenty-minute conversation in the host country language. While a number of worthwhile attributes can be cited for a procedure of this nature -- particularly its

high face validity as a test of active communication in the language -- the FSI interview technique cannot, or can only with difficulty, fill many of the requirements of an evaluation program intended to serve the various purposes listed above. A number of shortcomings in the FSI procedure from the viewpoint of its application to the Peace Corps situation can be briefly described.

The FSI rating appears to be relatively insensitive in the lower range of student competence. ETS staff members who have administered FSI-type interviews to trainees at the time of staging uniformly report a wide variation in language background (number of prior courses in the language, travel or study abroad) for trainees receiving FSI scores in about the 0 to 1 or 1+ range. In order to obtain more accurate baseline information about trainees' competence on entry into the Peace Corps program, more detailed and sensitive testing instruments seem needed. It should also be emphasized that accurate at-entry evaluation of language competence for various training groups would be an important prerequisite to making valid comparisons of training effectiveness across different training sites or curricula.

1. Listening comprehension. The nature of the interview is such that listening comprehension is tested only indirectly. It is always possible that the trainee's listening proficiency as such could be quite high, but that limitations in his ability to speak the language would prevent him from responding readily to questions or conversational

leads which he understood perfectly well. It is also the case that many listening comprehension situations that would be encountered in host country service (such as reception of radio broadcasts, telephone conversations, discussions among several people) cannot easily be presented in a face-to-face interview. At the higher levels of competence, where it would be useful, for example, to present very fast and/or colloquial or highly dramatic conversation, the realities of the interview situation are such that this type of testing is only rarely attempted, and certainly not on a consistent basis.

2. Extent of spoken vocabulary. The FSI interview does of course permit an estimation of general vocabulary level, but the specific vocabulary areas at issue in any given interview are largely dependent on the particular paths that the conversation takes. Although skilled interviewers attempt to cover certain general vocabulary content in the course of the interview, there can be substantial trainee-to-trainee variation in the type and level of vocabulary involved in or implied by the conversational topics treated. To the extent that the conversation deals with areas of vocabulary strength for a particular trainee (through his own interest, prior acquaintance with certain vocabulary domains) the entire interview is facilitated; the converse is true when the bulk of the

conversation happens to involve areas of experience having an unfamiliar vocabulary. For any given level of grammatical mastery or general fluency, the vocabulary implications of the topics that happen to be discussed may result in a substantial over- and under-rating for these other non-lexical aspects of the trainee's performance.

3. Command of spoken grammar. The FSI interview offers little opportunity to evaluate in a consistent and objective manner the trainee's command or lack of command of a wide number of different grammatical structures. Experienced FSI interviewers usually attempt to elicit the use of various verb tenses, and to a more limited degree, different persons beyond the first person singular; however, the bulk of the conversation is typically spent either in present- or simple past-tense discourse, with the trainee speaking in the "I" form almost exclusively. Further, the structure of the interview is such that the trainee is only rarely if ever obliged to ask questions of the interlocuter, that is, to use interrogative forms of the language. This last ability would be of particular importance to the volunteer, since he would be expected to do a substantial amount of "questioning" in the field. Other tenses or modes, such as indirect discourse, are virtually never at formal issue during the interview, and if they

appear at all, do so on a sporadic basis differing for each trainee and interview situation.

In summary of the above observations, the global and largely unsystematic nature of the face-to-face interviews currently employed as the major type of Peace Corps language evaluation are by their general nature unsuited to the close, diagnostic type of language measurement called for in many areas of concern to the Peace Corps, especially at the training-site or individual-trainee level. This is not to suggest that face-to-face interview techniques are wholly without merit; on the contrary, they are useful as a highly "visible" demonstration of the trainee's ability to sustain a conversation in the host country language. But due to their largely unstructured character they can afford little diagnostically useful information about the linguistic shortcomings or successes of particular trainees or language programs.

Suggestions for Changes in Language Evaluation Program

We would like to suggest that certain rather substantial changes be made in the nature and operation of the language evaluation program in an attempt to provide additional and more accurate measurement of PCT/PCV proficiency, for the goal purposes already described.

Such a program would make possible the collection of detailed information about the language proficiency of each individual trainee at important stages of his Peace Corps career.

While the individual PCT/PCV would be the basic "unit of measurement" in the evaluation program, it is of course to be understood that information obtained at the individual-trainee level could easily be combined and analyzed in terms of larger groups: the classroom, the training site, curriculum or program groupings across sites, all trainees or volunteers in a particular language. It should also be stated at this point that the development of effective language evaluation procedures would involve not only the consistent use of certain test instruments but also the routine collection of important language-related information such as prior study of or other exposure to the language, the nature of the language program at the training site, details of the host country job and other in-field experiences as they would affect the development of language skills. It is in drawing comparisons between such background or experience variables and test performance at various stages that the most valid and useful information about the operation of the language training program can be obtained.

Desirable Specifications for Language Evaluation Program

There are several general specifications which a comprehensive testing program for Peace Corps purposes should meet; some of the more important considerations are listed below:

1. The operational principles and basic format of the testing instruments should be such that similar instruments can be validly and straightforwardly developed in any of

the Peace Corps languages. Although the new testing procedures would initially be developed and used on only one or two of the higher-volume languages (tentatively, French or Spanish), it should be determined from the outset that the same general types of tests can readily be produced in other languages.

2. Since the primary impetus of Peace Corps language instruction is towards the development of listening comprehension and speaking proficiency, these two aspects of language command should receive primary attention in the test development program. Additional supplementary testing procedures for reading comprehension and writing ability may be developed at a later time for use in programs where these skills are applicable, but these would be of a second order of priority to the development of listening and speaking measures.

3. The tests should be so designed as to economize the time of both the PCT/PCV and the test administrator or other Peace Corps staff, consistent, of course, with valid measurement principles, which dictate certain minimum test components and overall testing lengths.

4. The tests should allow for valid and reliable administration under actual Peace Corps training-site and in-field conditions. In this regard, procedures which require elaborate equipment or which are in other respects complicated to administer would be avoided in favor of tech-

niques which would permit easy and straightforward administration by relatively untrained persons.

5. The tests should encompass a wide range of student proficiency, so that a single instrument could be used to test all students in a given training program or (using alternate forms) the same students at different points in their Peace Corps training and service. To the extent that a single test or test battery can be used to measure the entire range of proficiency usually encountered in the Peace Corps program, both test administration and test interpretation and use can be facilitated.

6. The tests should provide, in addition to summary numbers indicating overall competence, more detailed feedback to the trainee/volunteer and Peace Corps staff regarding language areas of strength or weakness, both on an individual and group basis.

Consideration of the above desired specifications has led to a tentative outline for a language testing program as described below. General test administration factors are discussed first, followed by a closer description of the proposed test instruments.

Description of Suggested Testing Program

It is intended to specify and develop language tests which can be administered at the testing sites by regular Peace Corps staff in these areas. However, during the developmental phase of the program we would like to adopt the position that testing at the different sites be carried out by a test administrator

provided by the ETS staff who would be responsible for:

1. Carrying or shipping the required test materials to the site;
2. Coordinating the scheduling of the testing with appropriate on-site personnel;
3. Explaining both the procedures and the underlying purposes of the testing program to on-site personnel and the trainees themselves;
4. Administering the tests and other instruments (such as any associated questionnaire);
5. Scoring the tests and informing the staff and trainees of the general and--where applicable--detailed results of the testing;
6. Returning test scores and associated data to ETS for inclusion in its data files and for appropriate reporting and research purposes.

There are several reasons for urging this general procedure during the developmental phase. On-site staff would for the near future at least be largely or completely unfamiliar with the testing program and as such would be expected to receive a shipment of testing materials and administration procedures with relatively little enthusiasm. A test administor sent from outside the site could serve as an important face-to-face information source about the testing program, and could probably exercise a greater positive influence for the program than any number of memoranda or other printed descriptions. Finally, in the early stages of test development a certain number of administration problems would be expected;

a trained administrator at the site would be in a good position to take corrective measures and also to make definite note of problems or irregularities which should be taken into account in refining the tests or administration procedures.

A typical on-site test administration might take place more or less as follows: The test administrator would reach the site in the afternoon or early evening and hold a 15-20 minute conversational session with the assembled students and any interested staff. The purpose of this meeting would be for the administrator to introduce himself, explain the reasons for the testing, and describe the procedures that would be followed. It is anticipated that an explanatory leaflet would also be distributed at this time, together with any questionnaire material which the PCT/PCV would be asked to fill out on his own following the meeting.

After the general discussion, a group test of listening comprehension would be administered. This would be an objective, multiple-choice test lasting approximately 20 minutes. The administrator would play a tape recording giving the spoken question or other stimuli, while the students would see panels of pictures or English options* in their test books and mark their answers on separate answer sheets. Tape recorded (rather than personally spoken) stimuli would be used for convenience and for uniformity of administration on a test-to-test basis. The spoken materials would range in difficulty from simple sentences delivered at a rather

* The use of English for printed answer options is based on the notion that students in a training program primarily devoted to listening and speaking activities should not be expected to be competent in reading the target language. While it is probably true that extensive use of English should be avoided in the classroom situation, we feel that this is not an important factor in the limited testing situation, and that the measurement advantages to be gained far outweigh any of the assumed drawbacks in using English options in the tests.

slow pace up through longer, more lexically and syntactically complex passages. Later sections of the test would include portions of typical radio broadcasts and telephone converations.

It is probable that certain portions of the listening comprehension test would be inappropriate for a given student, that is, either too easy or too difficult. We consider, however, that a few minutes of test inappropriateness at the beginning or end of the test would not be disconcerting to the student, particularly if he were informed in advance that the listening test (and other tests he would be taking) were deliberately of a very broad range.

Scoring of the listening test would be done on-site by the test administrator, permitting "real-time" feedback of the score information; the answer sheets themselves would be returned to ETS for item analysis and general research purposes.

The second and final phase of the testing would be carried out the following day on an individual-student basis: Students would come to the testing room in alphabetical order at 20-minute intervals. Approximately the first 5 minutes of this period would be spent in general but guided conversation between the student and administrator. Although this conversation would have the superficial appearance of an abbreviated FSI interview, it would differ from that procedure in two important respects. First, the conversation would be much more carefully structured in that the administrator would follow a specific protocol of questions-to-be-asked. This does not imply that there would be a rigid and strictly similar interview for each student; rather, the administrator would

have in mind (or on paper) sets of alternative possible questions or topics at a number of difficulty levels: these questions on topics would be alternated more or less randomly across students, with the overall effect one of reasonably spontaneous conversation.

A second major departure from the FSI procedure would be that only one basic aspect of the student's performance -- "general communicative fluency" -- would be evaluated. Accuracy of pronunciation, depth and extent of vocabulary, or knowledge of particular syntactic structures would <u>not</u> be at issue (these aspects would be evaluated separately as described below); rather, the student's ability to "get his message across" would be the primary consideration. Student performance in this respect could range from a very low category (considerable pausing obviously due to groping for appropriate means of expression; ambiguous or misleading information usually conveyed) to a very high one (near-native ease in conveying ideas, any potential blocks in fluency avoided through paraphrasing; interviewer never in doubt as to student's intended meaning).

It is of course obvious that the proposed short conversational interview is necessarily somewhat subjective in character and further, that the general "fluency" at issue in the interview is to a large extent dependent on developed competence in pronunciation, vocabulary, and syntax, to be later measured separately. Nonetheless, the preservation of a face-to-face conversation, albeit quite condensed in length from the regular FSI interview, is considered important for motivational reasons (after all, one of the most important goals of Peace Corps language training is to make it pos-

sible for the student to "talk to people" in the host country language); further, it is useful for research purposes to secure at least some measure of overall fluency as judged on a somewhat intuitive, face-to-face basis: correlations between a general fluency score of this type and other component aspects of the student's performance (pronunciation, lexicon, syntax) should provide a certain amount of insight about the relative contribution of these aspects to overall communicative ability.

An appropriate scoring scale for the conversational interview portion has not yet been determined, but this can be done quite easily following a number of trial administrations of a condensed interview. Some considerations in the development of the scale are: (1) the need to define a scale which is <u>not</u> congruent with the FSI scale nor readily convertible to this scale, and (2) the need to specify a large enough number of score categories to represent adequately the range of fluency encountered, without exceeding the number of discriminations that could reliably be made by the typical rater. It appears at present that a scale running from 1 to 7 would prove both sufficiently reliable and discriminating;* direct comparisons between a 1-7 scale and the regular FSI scale should be difficult enough to discourage any arbitrary and uninformed equating of the two tests. In this connection, it would

* The FSI interview uses eleven score categories (0 to 5, including plusses) of which only nine or ten are usually at issue (0 or 0+ to 4+). The 7-maximum score categories for the condensed interview would reflect the need to coarsen the new categories somewhat in keeping with the reduced testing time.

be important for ETS to emphasize to users of the test results that the condensed interview is not a FSI rating and does not of itself represent a reliable and definitive statement about the student's competence in the language: the general fluency score would, rather, be interpreted along with four other measures (listening comprehension, pronunciation, spoken lexicon, spoken syntax) in arriving at an overall picture of the student's language proficiency at a given point in time.

Immediately after the face-to-face conversation, the administrator would begin a rather highly automated test lasting not more than 10 to 15 minutes and evaluating the accuracy of the student's pronunciation, the breadth and depth of his active (speaking) vocabulary, and the extent of his structural command of the spoken language. Each of these skill aspects would be tested directly and insofar as possible independently of the other aspects. The general technique in all three cases would be for the student to look at pictures or English sentences printed in a test booklet and to make spoken responses based on the printed stimuli.

In the pronunciation section, the student would look at pictures representing extremely common objects for which virtually all students would be expected to know the host-country word, and would be asked to name the objects with particular attention to the accuracy of pronunciation. The spoken words represented by the pictures would be carefully chosen to embody important single sounds in the host-country language; the student's mastery of these sounds would be determined on a right-wrong basis by the test administrator using a keyed answer form on which

he would score the student's response immediately after it is given. The initial series of pictures in the test would embody sounds whose accurate pronunciation is necessary for unambiguous communication (phonemic criterion); later pictures would check the phonetic accuracy of certain sounds whose mispronunciation is not usually crucial to understanding but which have a bearing on the overall "foreignness" of the student's speech. A final portion of the pronunciation section would present very simple English sentences which the student would render aloud in the host-country language: this section would check the accuracy of intonation contours, liaison where appropriate, and other suprasegmental features of the student's pronunciation. Scoring would again be on a right-wrong basis and would be carried out by the administrator in the course of the test itself.

The second major section of the entire test would evaluate the extent and depth of speaking vocabulary: the student would see in his test booklet a rather large number of pictured objects or English words. In each case he would attempt to say aloud the foreign-language equivalent for the object or word. Unlike the situation for the pronunciation and grammatical control portions of the test (where the vocabulary aspects are deliberately held at a very elementary level in favor of testing <u>other</u> skill aspects), the vocabulary tested in this portion would range from extremely common, everyday terms up to fairly specialized vocabulary in a number of areas (though not so specialized as to be beyond the experience of the average native

speaker). Scoring would again be on a right-wrong basis and would be carried out at the time of testing using a marking form showing the acceptable responses and having a simple provision for checking each of the responses as correct, incorrect, or not attempted.

The final section of the test would evaluate the accuracy and extent of the student's command of the spoken structure of the language. The technique would again be that of presenting printed stimuli in English which the student would render aloud in the host-country language. The stimuli in English would be in the form of sentences or questions and would begin very simply (for example, "He is here") ranging upwards to much more syntactically complex utterances (such as "He would have gone"). Tested in this section would be the command of verb forms and tenses, comparatives of size, temporal expressions, pronoun use and placement, and other aspects usually included for that language under the general category of "syntax." Although the structural complexity of the stimulus sentences would show a progressive increase, the vocabulary in which the sentences are expressed would remain simple throughout to minimize lexicon as a factor in the student's responses. The structural control section would, as in the other parts of the test, be scored on a right-wrong basis simultaneously with test administration.

Test Information Feedback to Student and Language Staff

Following each individual interview, the test administrator would have in his possession the student's overall scores for

each of the several sections of the test battery: listening comprehension, general communication, pronunciation accuracy, vocabulary mastery, grammatical mastery. For the last three aspects, he would in addition have detailed information on the student's responses to the individual test questions. Theoretically, at least, it would be possible for the administrator simply to give a "carbon copy" of the detailed results to the student and the language staff. Although this would provide the maximum possible feedback, it is not considered desirable for reasons of test security to give the student or language staff facsimile copies which would show the specific questions asked (that is, to reproduce the pictures, printed words, or sentences actually used in the test).*

A compromise solution, which is considered to provide students and training staff a reasonably comprehensive and useful indication of strengths and weaknesses while safeguarding the details of particular test forms, would involve giving total test scores for each section, together with a description of the areas in which the student has or lacks mastery. Thus, for the pronunciation part of the test, the student would not learn that a particular word or phrase was rendered correctly or incorrectly, but rather, that his control of certain classes of sounds or sound patterns was acceptable or deficient. For

* "Pulbication" of the test in this manner would require that a new and different test be provided for each student--obviously an impossible ocndition.

the vocabulary section, the lexical domains of strength or weakness would be identified (such as: basic formulas of politeness or greeting, vocabulary of one's biography or general personal description, terms appropriate to the mechanics of travel, food buying, etc.). For the structural control section, the grammatical areas of mastery or lack of mastery would be indicated.

The above feed back procedure would appear to pose a formidable task for the test administrator. We feel, however, that it should be possible to automate the test administration, scoring, and feedback procedure so that the administrator would have only to make check marks or similar indications on a single scoring form: these notations would be automatically transferred to and converted to an appropriate "feedback format." The suggested technique would involve the use of NCR ("no carbon required") forms in sets of three sheets: the top sheet--seen and used by the test administrator--would show the detailed test stimuli (as contained in the test booklet) as well as the anticipated correct responses or other scoring guides. The second and third sheets, distributed to the student and language staff respectively, would show only the total scores and checks or other marks indicating success or lack of success in the various categories of performance discussed above. Necessary interpretive information would also be printed directly on the "feedback" copies of the scoring form.

Suggested Points for Test Administration During Training/ In-Service Sequence

We have attempted to define those points in the Peace Corps Training/in-service sequence at which language achievement tests of the type discussed could most profitably be administered. Two occasions have been identified as being of primary importance, staging, and completion of formal language training. A third testing point, for which we would suggest implementation in particular language/project combinations on an investigatory basis, would be following about four to six months of host-country service.

1. Staging For purposes of staging testing, it should be possible to separate all Peace Corps training programs into two major categories: those for which a reasonable proportion of the entering trainees would be expected to have had formal training or other prior exposure to the host country language (essentially French, Spanish, and possibly Portuguese); and languages for which virtually no prior contact would be anticipated. For training projects in the second category, achievement testing at staging would be inappropriate and time-wasting since all trainees would be assumed to be starting from "zero" achievement. Aptitude testing, on the other hand (using an instrument such as the Pimsleur Language Aptitude Test or the Carroll-Sapon Modern Language Aptitude Test) could provide a useful means for assigning trainees to slower or faster classes. Indeed, this is probably the only

basis on which students in these languages could be sectioned for training with any degree of validity or practical success.

For training programs in the first category, achievement testing at the time of staging would, however, be an important factor: test scores on the achievement battery would permit the assignment of trainees to class sections on the basis of demonstrated competence; these same tests would provide the specific baseline data on individual performance necessary for later comparisons of training effectiveness (that is, individual or group language improvement over the course of training would be represented by the differences in test performance on-entry and at the conclusion of the training program).

Aptitude test scores for "common-language" groups at staging would be useful for those students having no prior background in the language (in the same way that they would be useful for all students in "unknown language" programs). Aptitude tests for trainees having some background in the language would be important for research purposes, since they would provide a statistical control for "language aptitude"; aptitude scores could also usefully supplement information obtained from the achievement scores for purposes of class assignment.

ETS staff would be glad to administer an aptitude test at staging as part of the testing program, and to score these tests for immediate use in class placement.

2. End-of-Training. End-of-training achievement testing would provide an indication of the maximum trainee language proficiency at the time of entry into host country service. For trainees with some prior knowledge of the language, comparisons of end-of-training test results to similar data obtained at staging would indicate what portion of the student's proficiency could be attributed to Peace Corps instruction; for trainees in the uncommon languages, end-of-training performance would presumably be solely attributable to the training program.

Since the achievement tests would be designed to evaluate control of important language features independently of particular curricula or teaching methods, it should be possible to make valid end-of-training comparisons of language teaching success for different types of training programs or different training sites. For a given site or instructional method, chronological comparisons across a number of projects would indicate any significant trends in the overall quality of instruction.

3. In-field. Although a language achievement test battery would be somewhat more expensive and difficult to administer in the field, at least a limited amount of such testing is recommended to provide insight into the type and extent of language improvement that could be expected to take place as a consequence of the volunteer's normal interaction with the host-country speech community. For this purpose, testing in only a few "research-designated" programs or projects should provide sufficient data. An important consideration, of course, would be to test in the

field only volunteers who had previously been exposed to a full achievement testing program during training.

The suggestion that in-field testing should be done at about 4-6 months of host country service is only provisional, and reflects our appraisal of the earliest point in serviee at which tangible gains in language performance could be anticipated.* However, it would be useful (and probably necessary due to travel and scheduling complexities) to broaden this range somewhat on either end. The exact duration of volunteers' in-field service would of course be taken into account in interpreting test results.

Achievement Testing on Other Occasions

FSI-type testing has in the past been carried out at various points other than or in addition to the three discussed above; this is especially true for "mid-training" and at the end of host country service. Since it is hoped that course-of-training tests keyed to the curriculum of a particular training program can eventually be developed (as described in the following section), we see no substantial benefit in planning for or administering external achievement tests at a mid-training point.

Although end-of-service data on volunteer language proficiency would be of general interest, we feel that for research and

*Carroll (1966, A Parametric Study of Language Training in the Peace Corps) cites a range of 0-11 months during which volunteers reported lack of proficiency in various language skills. For those volunteers who were considered "non-qualified" in the language on entry into the field, average durations of reported language difficulties were concentrated at about 4-6 months, suggesting a general "point of improvement" at this time.

program development purposes such testing would simply be too late in the volunteer's service, and would provide much less useful information than would a host country administration conducted substantially earlier.

Development of Course-of-Training Evaluation Procedures

Observations and discussion with language staff at various training centers indicates that in most instances little or no testing of foreign language mastery occurs other than that carried out by ETS staff during on-site visits.

Where all training in a specific language is confined to a specific curriculum, the development by ETS -- with the advice and collaboration of appropriate PC staff -- of measurement materials suitable for assessing the trainees' foreign language attainment at various points during training would be considered feasible. Such measurement materials would, necessarily, be designed for use with the curriculum specified, and would provide language trainers with objective evidence of group and individual mastery or non-mastery of specific linguistic points as they were presented in the course of training. Such instruments would help trainers to adapt their techniques to the various groups being taught by revealing areas in need of extended drill, re-teaching or, conversely, by indicating mastery of a particular phenomenon and allowing the instructor to accelerate.

Often, however, for specific languages, and in particular those having high enrollments, a variety of teaching materials

and approaches are used. The lack of uniformity in materials or curricula makes extremely difficult the preparation of "course-of-training" test materials appropriate to all training programs in a specific language.

Where different programs exist for a given language, it would nonetheless be possible for ETS to provide orientation in testing and measurement to language trainers on site, in conjunction with its regular test administration visits. When appropriate, ETS staff members could extend their on-site visit beyond the time required for testing to conduct one-day training sessions in foreign language testing. These sessions would provide language instructors with information concerning means for evaluating trainee progress during the course of the training program, discovering weaknesses in instruction, and assisting attainment in specific skills. Appropriate materials would be assembled in the form of a "kit" that could be provided to the language training staff taking part in the orientation session as a resource for future reference.

An outline of topics that would be treated in a typical orientation session is shown below:

I. Purposes and Goals of Foreign Language Testing
   - A. to assess individual performance
   - B. to assess group performance
   - C. to discover weaknesses in the instructional program

II. Testing Techniques Language Skills
   - A. listening
   - B. *reading
   - C. *writing

*where appropriate to the program

D. "communication"

III. Ongoing Evaluation Procedures

A. informal classroom evaluation

B. course-of-training tests, examinations

IV. External-to-Program Tests of Overall Language proficiency

A. their general nature

B. their relationship to course-of-training measurement

Concluding Remarks

The language testing staff at ETS is highly interested in the problems and opportunities for effective language evaluation as embodied in the Peace Corps program. We would be very interested in working with appropriate Peace Corps personnel in defining and assisting in the development of a comprehensive program in this area, and toward this end, the preceding draft proposal is submitted.